International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03690
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A SURVEY PAPER ON PREDICTIVE ANALYTICS IN HEALTHCARE: A MULTI-DISEASE DIAGNOSIS SYSTEM FOR EARLY RISK ASSESSMENT

Mrs. J Bhargavi*[1], Garisapati Sai Uday*[2], Nallamasa Uday Kiran*[3], MD Aquibuddin*[4]
*[1]Assistant Professor of Department of CSE (AI & ML) of ACE Engineering College, Hyderabad, India.
[*2,3,4] Students of Department CSE (AI & ML) of ACE Engineering College, Hyderabad, India.

ABSTRACT

The Multiple Disease Prediction System is a machine learning-based tool designed to predict diseases based on symptoms provided by users. This system focuses on detecting Diabetes, Heart Disease and Parkinson's Disease which collectively impact a significant portion of the global population. Using Support Vector Machine and the Logistic Regression algorithms, the system enhances disease prediction accuracy and efficiency. Globally 10% of adults suffer from Diabetes according to the International Diabetes Federation (IDF), 32% of deaths worldwide are caused by Heart-Disease, making it the leading cause of mortality as per the World Health Organization (WHO),1% of people over 60 are diagnosed with Parkinson's Disease with prevalence increasing with age (Global Burden of Disease Study). The proposed system allows users to input symptoms through a web-based interface and instantly receive a diagnosis with probability scores. This minimizes the need for multiple medical consultations, making disease detection faster and more accessible.This project serves as a one-stop solution for users seeking preliminary medical diagnoses, ultimately contributing to better healthcare management and early disease intervention.

INTRODUCTION

Advancements in predictive analytics and machine learning are revolutionizing modern healthcare by enabling early detection, risk assessment, and personalized treatment. This project introduces a multi-disease diagnosis system designed to predict the likelihood of Diabetes, Heart Disease, and Parkinson's Disease based on user-provided symptoms and clinical data. These diseases are among the leading global health concerns—Diabetes affects approximately 10% of adults worldwide, Heart Disease remains the primary cause of mortality, and Parkinson's Disease impacts a growing number of individuals over the age of 60. Early diagnosis is essential for improving outcomes, reducing healthcare costs, and enhancing patient quality of life. The proposed system utilizes machine learning algorithms, specifically Support Vector Machine (SVM) and Logistic Regression, to deliver accurate, real-time predictions. Through a user-friendly web interface, individuals can input relevant health information and receive immediate diagnostic feedback with associated probability scores. By minimizing the need for repeated clinical visits and supporting proactive medical decision-making, this system improves accessibility to preliminary healthcare assessments. It also serves as a scalable and efficient tool for early intervention, ultimately contributing to better health management and disease prevention. Integrating predictive analytics into routine diagnostics represents a vital step toward smarter, data-driven, and more inclusive healthcare delivery systems.

BACKGROUND OF THE PROJECT

The global healthcare landscape faces significant challenges due to the rising prevalence of chronic and age-related diseases such as Diabetes, Heart Disease, and Parkinson's Disease. These conditions often remain undiagnosed until they reach advanced stages, making early intervention critical for better outcomes. Traditional diagnostic methods, which involve in-person consultations and laboratory tests, can be time-consuming, expensive, and inaccessible in some regions. In contrast, advancements in machine learning and predictive analytics offer promising solutions for early disease detection by analyzing large datasets and identifying hidden patterns linked to various health conditions. Multi-disease diagnosis systems leverage these technologies to simultaneously assess the risk of multiple diseases based on user-inputted symptoms and demographic information, enabling quicker and more accurate preliminary diagnoses. This project aims to develop a machine learning–based system using algorithms such as Support Vector Machine (SVM) and Logistic Regression, designed to provide real-time risk assessments for Diabetes, Heart Disease, and Parkinson's Disease. By offering an accessible and efficient tool for early detection, this system seeks to support proactive healthcare management, reduce diagnostic delays, and ultimately empower individuals to take preventive measures for better health outcomes.

LITERATURE REVIEW

1.Title: A Review on Machine Learning Algorithms for the Diagnosis of Multiple Diseases
Authors: Dua D, Singh A, and Bansal M (2021)

This review explores how various machine learning (ML) algorithms—such as decision trees, SVM, k-NN, and ensemble methods—are applied in multi-disease diagnosis. It highlights the potential of hybrid models to increase diagnostic accuracy and efficiency, particularly for chronic and comorbid conditions.

2.Title: A Unified Framework for Early Detection of Multiple Diseases Using Deep Learning
Authors: Zhang Y, Lin X, and Huang T (2022)

The authors present a deep learning-based framework capable of simultaneously detecting early-stage symptoms of diseases such as diabetes, cardiovascular disorders, and kidney failure. The model uses patient data from electronic health records (EHRs) and emphasizes temporal modeling through LSTMs.

3.Title: Early Prediction of Chronic Diseases using Machine Learning Techniques
Authors: Kaur P, Sharma M, and Mittal M (2020)

This study focuses on the predictive power of machine learning techniques like logistic regression and random forest for chronic disease detection. The authors discuss data preprocessing, feature selection, and class imbalance as critical components in improving prediction accuracy.

4.Title: Multi-Disease Prediction using Ensemble Learning: An Empirical Study
Authors: Kumar R, Singh V, and Sharma A (2023)

The paper evaluates several ensemble learning methods (e.g., XGBoost, AdaBoost) on health datasets and demonstrates how they outperform single classifiers for simultaneous diagnosis of diseases like hypertension, diabetes, and heart disease.

5.Title: Predictive Models for Health Risk Assessment Using EHR Data
Authors: Chen M, Hao Y, and Li Y (2021)

This study utilizes real-world EHR data and proposes predictive models for assessing patient risk profiles. The work stresses the importance of data quality, missing data handling, and ethical considerations in building scalable healthcare applications.

6.Title: A Hybrid Model for Multi-Disease Diagnosis with Explainable AI
Authors: Patel R, Singh S, and Rajan S (2022)

The authors propose a hybrid system combining rule-based inference with ML models and SHAP for explainability. It enables clinicians to understand and trust predictions made for coexisting diseases, particularly in elderly populations.

7.Title: Deep Multi-Task Learning for Joint Disease Prediction
Authors: Li H, Wang F, and Chen T (2023)

This research introduces a multi-task learning approach to predict multiple diseases in parallel. The model learns shared representations to enhance performance across tasks while accounting for disease-specific patterns.

8.Title: Big Data Analytics in Healthcare: Challenges and Opportunities
Authors: Ristevski B and Chen M (2018)

While not specific to multi-disease systems, this foundational work discusses how big data technologies can enable advanced predictive analytics in healthcare, particularly by integrating wearable, clinical, and genomic data..

9.Title: Early Risk Assessment Models Using Federated Learning in Healthcare
Authors: Nguyen D C, Ding M, and Niyato D (2022)

This paper investigates privacy-preserving FL models for early disease risk detection across multiple hospitals. It focuses on building robust and generalizable predictors without centralizing sensitive patient data.

10.Title: Survey of AI Applications in Multi-Disease Prediction
Authors: Ahmed M, Hasan M, and Islam M R (2020)

This survey compiles a broad overview of AI applications in diagnosing multiple diseases simultaneously. It outlines recent trends, dataset challenges, model comparisons, and future directions in the domain.

RESEARCH GAPS IN EXISTING SYSTEMS

Despite considerable progress in applying machine learning to healthcare diagnostics, current systems for multi-disease prediction and early risk assessment still exhibit several key limitations:

1. **Lack of Standardized, Large-Scale Multi-Disease Datasets**
Most existing models are trained and validated on disease-specific datasets or small, institution-specific samples. There is a significant lack of publicly available, high-quality, and multi-label datasets that represent diverse populations and comorbid conditions. This limits the generalizability and robustness of predictive models across different clinical environments.

2. **Limited Integration of Temporal and Longitudinal Data**
Many predictive systems rely on static snapshots of patient data rather an longitudinal health records that reflect disease progression over time. The failure to fully utilize temporal patterns in patient history weakens the models' ability to detect early warning signs or assess future risks effectively.

3. **Insufficient Explainability and Clinical Interpretability**
While high-performing black-box models like deep neural networks are increasingly used, their lack of transparency remains a major concern for real-world adoption. Clinicians require interpretable outputs that support trust, accountability, and informed decision-making. Few systems incorporate explainable AI (XAI) frameworks to bridge this gap.

4. **Underrepresentation of Real-Time and Resource-Constrained Environments**
Many models assume ideal computational resources and do not account for deployment on low-power or real-time healthcare systems, such as mobile health apps or embedded clinical decision support tools. This results in practical limitations when applying predictive analytics in under-resourced or remote settings.

5. **Limited Personalization and Adaptability**
Current multi-disease prediction systems often take a one-size-fits-all approach. They do not adequately adapt to individual patient characteristics such as genetics, lifestyle, or socio-demographic factors. This restricts their effectiveness in delivering personalized risk assessments and recommendations.

PROPOSED SYSTEM

The proposed system presents an intelligent, scalable, and clinically relevant framework for early risk assessment and multi-disease diagnosis using predictive analytics. It is designed to assist healthcare providers in identifying and managing the early signs of multiple conditions—such as diabetes, cardiovascular diseases, hypertension, and kidney disorders—by leveraging heterogeneous patient data sources. The system begins with a data integration layer that collects and preprocesses structured and unstructured health data from electronic health records (EHRs), wearable devices, lab test results, and patient-reported outcomes. Robust data cleaning, normalization, and anonymization techniques ensure quality and privacy compliance. At the heart of the system lies a multi-label classification engine built on a hybrid of traditional machine learning algorithms and deep learning models. This allows simultaneous detection of multiple diseases and accommodates complex interactions among comorbid conditions. To support personalized diagnosis, a risk stratification module analyzes demographic, genetic, behavioral, and historical medical data to assign individual risk scores and classify patients into actionable risk tiers. Enhancing the system’s transparency and clinical reliability, an Explainable AI (XAI) layer is incorporated using techniques such as SHAP and LIME to provide interpretable insights into how specific predictions are made. These explanations help clinicians trust the model’s outputs and integrate them into decision-making processes. The system includes a real-time decision support interface that offers intuitive visualizations, health monitoring dashboards, and early alert mechanisms, accessible via both desktop and mobile platforms. Additionally, a continuous learning mechanism enables the model to evolve over time based on new clinical outcomes and patient feedback, thereby improving predictive performance and adaptability.

CONCLUSION

In conclusion, the proposed predictive analytics system offers a robust and scalable solution for early detection and risk assessment of multiple diseases, addressing critical challenges in current healthcare diagnostics. By integrating diverse health data sources and employing advanced machine learning techniques, the system enables simultaneous identification of comorbid conditions with high accuracy. Its emphasis on personalized risk profiling, real-time decision support, and explainable AI ensures that the model is both clinically useful and trustworthy. Furthermore, its adaptability to resource-constrained environments through model optimization makes it suitable for wide-scale deployment, including in underserved regions. Ultimately, this system has the potential to significantly enhance early diagnosis, improve patient outcomes, and support data-driven decision-making in modern healthcare**..** The framework's scalability, cost-effectiveness, and ability to function in both urban and resource-limited rural settings make it a viable solution for addressing healthcare disparities. With ongoing enhancements in data privacy, model interpretability, and integration with telemedicine platforms, the system holds significant promise in supporting early intervention, reducing healthcare burdens, and ultimately saving lives. Future extensions could include personalized treatment recommendation modules and integration with federated learning for privacy-preserving collaborative training across hospitals or regions.s

REFERENCES

[1] S. Dey, A. Ashour, and S. Chakraborty, "Machine learning techniques for diagnosis of diseases and health status monitoring," *Journal of Healthcare Engineering*, vol. 2021, Article ID 9925432, pp. 1–15, 2021.

[2] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019..

[3] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[4] S. Miotto, L. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities " *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018..

[5] K. Reddy, R. K. Reddy, and B. T. Rao, "A predictive model for early detection and classification of multiple chronic diseases using machine learning techniques," *Materials Today: Proceedings*, vol. 62, pp. 8474–8480, 2022.

[6] H. Haider, A. A. Khan, and M. Qamar, "An intelligent multi-disease diagnosis system using machine learning for healthcare," *Computers in Biology and Medicine*, vol. 147, Article ID 105781, 2022.

[7] N. A. Alzahrani, F. Alshammari, and M. Aljuaid, "Explainable AI for healthcare: A survey on trends and challenges," *Healthcare*, vol. 11, no. 3, pp. 1–21, 2023.

[8] S. Choudhury and S. Mazumder, "A deep learning-based framework for early detection of multiple diseases using EHR data," in *Proc. 2021 IEEE Int. Conf. Healthcare Informatics (ICHI)*, Victoria, BC, Canada, 2021, pp. 421–427.

[9] T. Shortliffe and J. J. Cimino, *Biomedical Informatics: Computer Applications in Health Care and*

Biomedicine, 5th ed. Cham, Switzerland: Springer, 2021.

[10] Y. Zhang, Y. Chen, J. Wang, and M. Zhang, “Federated learning for healthcare: Opportunities and challenges,” *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12820–12835, Aug. 2021.